Example. Let $X \sim \text{expo}(10)$, $Y = \log X$ $(= \log_e X)$. Then $y = \log x = h(x)$, $x = e^y = h^{-1}(y)$ and

$$\begin{aligned} f_Y(y) &= \left|\frac{dx}{dy}\right| f_X(x) = \left|\frac{dx}{dy}\right| \frac{1}{10} e^{-x/10} 1_{(0,\infty)}(x) = \left|\frac{de^y}{dy}\right| \frac{1}{10} e^{-e^y/10} 1_{(0,\infty)}(e^y) \\ &= \frac{e^y}{10} e^{-e^y/10} 1_{(-\infty,\infty)}(y). \end{aligned}$$

Example. Consider an absolutely continuous r.v. X and $Y = h(X) = X^2$. Note that h is decreasing on $A_1 = (-\infty, 0)$ and increasing on $A_2 = (0, \infty)$, i.e. h can be decomposed into invertible parts h_1 and h_2 defined on A_1 and A_2. Then

$$\begin{aligned} F_Y(y) &= P(Y \le y) = P(X^2 \le y) = P(-\sqrt{y} \le X \le \sqrt{y}) \\ &= F_X(\sqrt{y}) - F_X(-\sqrt{y}), \\ \Rightarrow \quad f_Y(y) &= \frac{1}{2\sqrt{y}} f_X(\sqrt{y}) - (-\frac{1}{2\sqrt{y}}) f_X(-\sqrt{y}) \\ &= |\frac{dh_2^{-1}(y)}{dy}| f_X\{h_2^{-1}(y)\} + |\frac{dh_1^{-1}(y)}{dy}| f_X\{h_1^{-1}(y)\}. \end{aligned}$$

[Theorem 2.1.8: consider k disjoint intervals $A_i = (a_i, b_i)$, $i = 1, \ldots, k$, where h is monotone, then $f_y(y) = \sum_{i=1}^k f_X\{h_i^{-1}(y)\} |\frac{d}{dy} h_i^{-1}(y)|$.]

Assume additionally that X is normally distributed, $X \sim N(0,1)$, $f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} = f_X(-x)$ (symmetrical). Then $Y = X^2$ has a χ^2 distribution:

$$f_Y(y) = 2 \cdot \frac{1}{2\sqrt{y}} f_X(\sqrt{y}) = \frac{1}{\sqrt{y}} f_X(\sqrt{y}) = \frac{1}{\sqrt{y}} \frac{1}{\sqrt{2\pi}} e^{-y/2} \qquad \text{for } y > 0.$$

Preliminaries. Assume you wish to simulate a r.v. Y with continuous cdf F_Y but your random number generator only generates $U \sim U(0,1)$ outcomes. Then you can still simulate Y by calculating transformations $F_Y^{-1}(U)$, provided F_Y^{-1} exists. This holds due to the following theorem.

Theorem. Suppose F is a cdf such that it is continuous and 1-1. If $U \sim U(0,1)$ and $Y = F^{-1}(U)$ then $Y \sim F$.

Proof. Use the fact that F is strictly increasing and that $F_U(u) = u$ for $u \in (0,1)$ to obtain $P(Y \le y) = P(F^{-1}(U) \le y) = P(U \le F(y)) = F(y)$. □

Example. To simulate $Y \sim \text{expo}(\beta)$, $\beta > 0$, take the transformation $F_Y^{-1}(u) = -\beta \log(1-u)$: $F_Y(y) = 1 - e^{-y/\beta} = u \iff 1 - u = e^{-y/\beta} \iff \log(1-u) = -y/\beta \iff y = -\beta \log(1-u)$.

2.2 Expectation

Example. Let $Y \sim \text{binomial}(n, p)$ (number of successes). Consider first SWR (sample size n) from a population of size N and with $M = pN$ "successes". There are N^n equally likely outcomes. What is the average value of the number of successes $Y(s)$ (or of $h(Y(s))$)?

The number of samples with $Y(s) = y$ $(y = 0, 1, \ldots, n)$ is $\binom{n}{y} M^y (N-M)^{n-y}$. Thus, the average of Y is

$$
\begin{aligned}
\frac{1}{N^n} \sum_{s \in \mathcal{S}} Y(s) &= \frac{1}{N^n} \sum_{y=0}^{n} \sum_{s: Y(s)=y} Y(s) = \frac{1}{N^n} \sum_{y=0}^{n} y \binom{n}{y} M^y (N-M)^{n-y} \\
&= \sum_{y=0}^{n} y \binom{n}{y} p^y (1-p)^{n-y} = \sum_{y=0}^{n} y f_Y(y).
\end{aligned}
$$

Likewise, with Y replaced by $h(Y)$, $\dfrac{1}{N^n} \sum_{s \in \mathcal{S}} h(Y(s)) = \sum_{y=0}^{n} h(y) f_Y(y)$. The averages depend only on the pmf of Y, not on $\mathcal{S}$, which suggests the following definition (for any pmf).

Definition. Let Y be a discrete random variable.

(i) If h is non-negative then the **expectation** (average, mean) of $h(Y)$ is $E(h(Y)) = \sum_y h(y) f_Y(y)$ (can be ∞).

(ii) More generally, if $E(|h(Y)|) < \infty$ then we define $E(h(Y)) = \sum_y h(y) f_Y(y)$.

Example (cont.) Consider $Y \sim \text{binomial}(n, p)$ and note that $y\binom{n}{y} = n\binom{n-1}{y-1}$. Therefore,

$$
\begin{aligned}
E(Y) &= \sum_{y=0}^{n} y \binom{n}{y} p^y (1-p)^{n-y} = np \sum_{y=1}^{n} \binom{n-1}{y-1} p^{y-1} (1-p)^{n-y} \\
&= np \sum_{x=0}^{n-1} \binom{n-1}{x} p^x (1-p)^{n-1-x} = np.
\end{aligned}
$$

Similarly, using $y(y-1)\binom{n}{y} = n(n-1)\binom{n-2}{y-2}$, one can show that $E\{Y(Y-1)\} = n(n-1)p^2$.

For $X(s) = 1_A(s)$ with $P(A) = p$ we have $X \sim \text{Bernoulli}(p) = \text{binomial}(1, p)$ and therefore $E(X) = p$. Directly: $E(1_A) = 0 \cdot P(X=0) + 1 \cdot P(X=1) = 1 \cdot P(A) = p$.

In order to define the expectation of a continuous r.v. X we can, for example, make it discrete: consider the r.v. Y with $P(Y = j/n) = f_X(j/n)\frac{1}{n}$. Then $E(X) \approx E(Y) = \sum_{j=-\infty}^{\infty} \frac{j}{n} f_X(\frac{j}{n}) \frac{1}{n} \approx \int_{-\infty}^{\infty} x f_X(x)\, dx$.

Definition. Let Y be a continuous r.v. with pdf f_Y and h a real-valued function. (i) For h non-negative the **expectation** (average, mean) of $h(Y)$ is $E\{h(Y)\} = \int_{-\infty}^{\infty} h(y) f_Y(y)\, dy$ (can be ∞). (ii) More generally, $E\{h(Y)\} = \int_{-\infty}^{\infty} h(y) f_Y(y)\, dy$ if $E(|h(y)|) < \infty$.

Example. Consider $X \sim \text{expo}(\beta)$, i.e. $F_X(x) = 1 - \exp(-x/\beta)$ for $x > 0$, and $h(x) = x^m$, $m = 1, 2, \ldots$ Now use $\Gamma(m+1) = \int_0^\infty y^m e^{-y} dy = m!$ to obtain

$$
\begin{aligned}
EX^m &= \int_0^\infty x^m f_X(x)\, dx = \int_0^\infty x^m e^{-x/\beta} \frac{1}{\beta}\, dx = \int_0^\infty (y\beta)^m e^{-y}\, dy \\
&= \beta^m \int_0^\infty y^m e^{-y}\, dy = \beta^m m!.
\end{aligned}
$$

Remark. The expectation is linear and monotone, which is easy to see since expectations are integrals. (See Theorem 2.2.5.)